

Estimating time: comparing the accuracy of estimation methods for interval timing

Article (Published Version)

Damsma, Atser, Schlichting, Nadine, van Rijn, Hedderik and Roseboom, Warrick (2021)
Estimating time: comparing the accuracy of estimation methods for interval timing. *Collabra Psychology*, 7 (1). a21422 1-15. ISSN 2474-7394

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/101697/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Cognitive Psychology

Estimating Time: Comparing the Accuracy of Estimation Methods for Interval Timing

Atser Damsma^{1, a}, Nadine Schlichting^{1, b}, Hedderik van Rijn¹, Warrick Roseboom²

¹ Department of Experimental Psychology, University of Groningen, The Netherlands; Research School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands, ² Department of Informatics, University of Sussex, Brighton, UK; Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

Keywords: time perception, interval estimation, method comparison, motor reproduction, spatial estimation, verbal estimation

<https://doi.org/10.1525/collabra.21422>

Collabra: Psychology

Vol. 7, Issue 1, 2021

In interval timing experiments, motor reproduction is the predominant method used when participants are asked to estimate an interval. However, it is unknown how its accuracy, precision and efficiency compare to alternative methods, such as indicating the duration by spatial estimation on a timeline. In two experiments, we compared different interval estimation methods. In the first experiment, participants were asked to reproduce an interval by means of motor reproduction, timeline estimation, or verbal estimation. We found that, on average, verbal estimates were more accurate and precise than line estimates and motor reproductions. However, we found a bias towards familiar whole second units when giving verbal estimates. Motor reproductions were more precise, but not more accurate than timeline estimates. In the second experiment, we used a more complex task: Participants were presented a stream of digits and one target letter and were subsequently asked to reproduce both the interval to target onset and the duration of the total stream by means of motor reproduction and timeline estimation. We found that motor reproductions were more accurate, but not more precise than timeline estimates. In both experiments, timeline estimates had the lowest reaction times. Overall, our results suggest that the transformation of time into space has only a relatively minor cost. In addition, they show that each estimation method comes with its own advantages, and that the choice of estimation method depends on choices in the experimental design: for example, when using durations with integer durations verbal estimates are superior, yet when testing long durations, motor reproductions are time intensive making timeline estimates a more sensible choice.

Introduction

In the growing research field on interval perception the number of ways to measure subjective time are seemingly growing, too. As a researcher, one has to decide whether a task is retro- or prospective (e.g., Block et al., 2018), in which modality intervals are presented (e.g., auditory or visually; Wearden et al., 2006), how exactly intervals are presented (e.g., filled or empty; Grondin, 1993), the paradigm used (e.g., temporal reproduction, production, bisection, or comparison; for a review, see Grondin, 2010; Wearden, 2016), and how responses are being collected (e.g., verbal or motor responses; e.g., Block et al., 2018; Mioni, 2018). While subjective (distortions of) time perception may be captured no matter which choice was made regarding the listed options, often neglected from this choice are the potential differences in cognitive strategy or what representation of time underlies a given task.

One prominent idea is that time is represented in spatial terms (for a review, see Bender & Beller, 2014). Indeed, visuospatial representations of time are reflected in how we think and communicate about time, and also in how we process and act on time (Bonato et al., 2012; Núñez & Cooperrider, 2013). For example, time-related notions in language are often spatialized: the future lies ahead of us, we are looking back at earlier times, or the vacation was too short. The latter notion, how we process and act on time, is reflected in the commonly found Spatial-Temporal Association of Response Codes (STEARC) effect (Conson et al., 2008; Fabbri et al., 2012, 2013; Ishihara et al., 2008; Vallesi et al., 2008; Vicario et al., 2008; Weger & Pratt, 2008). The STEARC effect describes a space-related representation of time and temporal magnitudes, such that *before/shorter* responses have a processing or response advantage when associated with the left side of space, and, vice versa, *after/longer* responses show the same advantages

^a Shared first authorship; a.damsma@rug.nl

^b Shared first authorship

when associated with the right side of space. This spatialization of time can also be observed in children as young as five years (Coull et al., 2018). Mental timeline theories in particular suggest that time is represented as a spatial linear axis that allows absolute (i.e., how long a stimulus lasted) and relative timing (i.e., temporal order; Bonato et al., 2012; Magnani & Musetti, 2017). The orientation of the timeline is heavily influenced by culture and experience, such as, for example, reading direction (e.g., English speakers, who read from left to right, map events on a timeline directed rightward, while Arabic speakers, who read from left to right, showed the reverse pattern; Boroditsky, 2001; Fuhrman & Boroditsky, 2010) or commonly used spatial metaphors to talk about time (e.g., Mandarin speakers use horizontal and vertical terms to temporally order events, while English speakers commonly use only horizontal terms; Boroditsky et al., 2011). A number of neurobiological and cognitive models even suggest that space and time share their neural representation (e.g., A Theory Of Magnitude (ATOM): Walsh, 2003, 2015; hippocampal time and space cells: Buzsáki & Llinás, 2017), emphasizing the intertwinedness of these two dimensions. ATOM, for example, is based on 1) behavioral findings showing a tight link between spatial (size, length) and temporal magnitudes, in that spatial magnitude influences the perception of temporal magnitudes in a “more is more” fashion (i.e., more spatial magnitude is more temporal magnitude, Cai et al., 2018; Cai & Connell, 2016; Casasanto & Boroditsky, 2008; Xuan et al., 2007); and 2) on neuroimaging studies revealing shared neural representations in the parietal cortex during the processing of spatial, numerical and temporal magnitudes (e.g., Bueti & Walsh, 2009; Dormal et al., 2012; Hayashi et al., 2013; Riemer et al., 2016). Adding to this theory, Coull & Droit-Volet (2018) highlight that explicit representations of time are not solely rooted in space but also in motor interactions with the world, which have a temporal and a spatial component. The authors offer a developmental approach of how we construct a representation of time by performing actions in space during childhood (see also Loeffler et al., 2018).

Assuming that time is indeed represented spatially or in an ATOM-like common magnitude system, an additional method to estimate intervals is the use of a timeline or visual analogue scale. While visuospatial estimation formats are commonly used in intentional binding studies (e.g., Haggard et al., 2002), to our knowledge, only few interval timing studies have made use of them (e.g., Damsma et al., 2018; Roseboom et al., 2019). Apart from the more conceptual question of how exactly time may be represented in the brain, there are practical issues regarding the implications for different response modes at hand, too: So far it has not been tested whether an explicit translation from time to space affects precision and/or accuracy of temporal estimates compared to other commonly used estimation methods. In two separate experiments we aimed to test the advantages and disadvantages of using different estimates of time, namely reproductions in the time dimension, estimates in the spatial dimension, or estimates in a symbolic form.

In Experiment 1 participants estimated intervals by either pressing a button (motor reproduction), clicking on a

timeline (timeline estimation), or giving a numerical estimate (verbal estimation). The to-be-estimated interval was a white square appearing and disappearing on a black screen. The results of Damsma et al. (2018) suggested that participants exhibit a response bias when using timeline estimations, seen in avoidance of clicking close to the end of the line or screen. To test whether this bias can be prevented participants performed one of two versions of this experiment: one in which the range of the timeline corresponded to the tested intervals, and one in which the timeline corresponded to intervals longer than the tested intervals. In other words, participants were either calibrated to the test durations or to slightly longer durations. When estimating an interval using a timeline or verbal estimates, participants can be more deliberate in their estimates (i.e., go back and forth in time) compared to motor reproductions, in which participants have only one chance to make an estimate. While intervals had a clear on- and offset and required no further processing steps in Experiment 1, we used a more complex temporal estimation task in Experiment 2. Participants saw a stream of digits and one target-letter and were subsequently asked to first estimate the onset of the target letter within the stream, and second to reproduce the duration of the complete stream by either motor reproductions or timeline estimations. Again, half of the participants were calibrated to the test durations, while the other half were calibrated to longer durations. In this more complex setup, participants did not only have to attend to and memorize one duration, but they had to attend to the content of the stream and memorize two durations. Timeline estimates may allow for relative timing (e.g., when did the target occur relative to the estimated offset), while motor reproductions require a strictly sequential order of interval reproductions. In both experiments we will compare accuracy (i.e., the estimations) and precision (i.e., the absolute error and the coefficient of variation (CV)) of temporal estimates. A common finding in temporal estimation tasks, especially in reproduction tasks, is that previously encountered intervals influence the perception of the current interval (also known as sequential context effects; for a review, see van Rijn, 2016). We will compare the magnitude of these context effects for temporal estimation methods and calibration conditions. If there is a cost to a potential spatial transformation, we expect that the timeline estimates show lower accuracy and/or precision than motor reproductions and verbal estimates. In addition, we expect that calibration with longer intervals may increase the accuracy of the estimates, especially for longer intervals, in the timeline estimation condition.

Experiment 1

Methods

Participants. Sixty healthy adults (20 male, mean age 22.65) participated in exchange for course credits or a financial compensation of €8. All participants had normal or corrected-to-normal vision. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen (identification number 17408-S-NE) was obtained before testing. Sample size was based on past research (e.g., Damsma et al., 2018; Schlichting et al., 2018),

no statistical *a priori* power analysis was conducted.

Experimental Design and Procedure. Participants were asked to perform a temporal estimation task using three different estimation methods: motor reproduction, timeline and verbal estimation. At the beginning of the experiment, participants were explicitly instructed to not keep track of time by, for example, counting or feet tapping. Stimuli were displayed on a 1920 × 1080 LED-based monitor screen (Iiyama ProLite G2773HS) with a refresh rate of 100 Hz.

The to-be-reproduced interval (equally spaced between 1 and 4 s in steps of 0.5 s) was presented at the beginning of the trial. Appearance of a white square (50 by 50 pixels) at the center of the screen marked the onset, and the disappearance of the square the offset of the interval. After a fixation period of 1 s participants were asked to estimate the previously perceived interval in one of three ways: a) timeline estimation: participants were asked to click on a timeline at the point where the interval ended (1 pixel on screen corresponded to 0.01 s), apart from a tick demarking the onset of the interval there were no further spatial/temporal indications (ticks) given; b) motor reproduction: the white square re-appeared, marking the onset of the reproduction, and participants were asked to press the spacebar to end the interval; c) verbal estimation: participants were asked to enter a numerical estimate in seconds with one decimal place. After the estimation participants received immediate feedback on each trial (practice and experimental trials) in form of a timeline (1 pixel on screen corresponded to 0.01 s). The feedback format was the same for all conditions to make the tasks as equal as possible. Two grey bars on top of the timeline depicted the on- and offset of the veridical interval, and two white bars below the timeline depicted participants estimates (i.e., both onset bars were always aligned). See Figure 1 for a schematic depiction of the experimental design. The experiment was run in Matlab R2014b (The MathWorks) using the Psychophysics Toolbox version 3.0.12 (Brainard, 1997) in Windows 10.

The experiment was divided into three blocks (i.e., one block for each estimation method) of 42 experimental trials (i.e., four trials per duration) each. Order of blocks, and thus estimation methods, was counterbalanced between participants. Before the start of each block, participants received instructions about the estimation method to be used in the upcoming block, and they performed 12 practice trials in order to get accustomed to the timeline and the estimation method before the start of the experimental trials. The order of trials was the same in each block, but varied between participants.

As a between-subjects manipulation, half of the participants were assigned to perform a calibrated version of the estimation tasks. In the calibrated version the training-trials consisted also of longer intervals than those in the test trials (1.0, 2.5, 5.0, and 6.0 s), while in the uncalibrated version training trials were chosen from the range of intervals of test trials (1.0, 2.0, 3.0, and 4.0 s). Importantly, this changed the length of the timeline in the feedback screen and also in the timeline estimation condition: in the calibrated version the timeline was longer, so that during test trials participants did not have to click as close towards the end of the line to estimate the longest duration

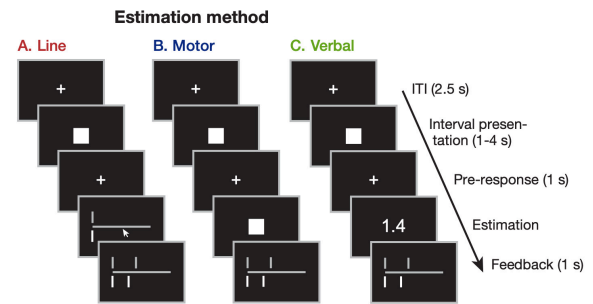


Figure 1. Trial procedure of Experiment 1. Participants performed a simple temporal estimation task, in which they had to estimate the duration of a square in three ways: A) clicking on a timeline (line estimation), B) pressing a key to indicate the estimated offset of the interval (motor reproduction), and C) typing a verbal estimate in seconds (e.g.: “1.4”; verbal estimation). Feedback was presented at the end of each trial.

as they had to in the uncalibrated version. In both the calibrated and uncalibrated condition, the timeline was presented centrally on the screen. The experiment files can be found at <https://osf.io/w38qg/>.

Analysis. All estimates shorter than 0.2 s and longer than 10 s (0.34% of the data) and all trials in which no estimates were provided (0.34% of the data) were excluded from analysis. The estimates were analyzed using Linear Mixed Models (LMMs) from the *lme4* package (Bates et al., 2015) in R (R Core Team, 2016). To compare the accuracy of the different conditions, we tested a model predicting estimates. In addition, we compared the precision of the conditions by testing models predicting the absolute error and the CV. Finally, we tested a model predicting reaction time. In each model, *duration* (i.e., the veridical duration of the interval), *estimation method* (motor, timeline or verbal) and *calibration condition* (uncalibrated or calibrated) and their interactions were sequentially added as fixed factors. Only fixed factors that significantly improved the model according to a likelihood ratio test were included in the final model. To assure the interpretability of significant interaction terms, the relevant main effects were also included in the model. In addition, the fixed factor *duration* was centered at 2.5 and *calibration condition* was recoded using effect coding (−0.5 and 0.5 for uncalibrated and calibrated, respectively), to make main effects of *duration* and *estimation method* easier to interpret. *Participant* was always included as a random intercept term. After establishing the final model, we sequentially added random slope terms, starting with the random slope that decreased the AIC value most. We tested whether the inclusion of the random slope term was warranted using likelihood ratio tests. Given the final model, we compared the three estimation methods with post-hoc contrasts using the *glht* function in the *multcomp* package in R (Hothorn et al., 2017).

Using LMMs enables us to test both intercept effects (e.g., if estimates in one condition are systematically over-estimated compared to another condition) and slope effects

(e.g., if there is a stronger pull towards the mean in one condition the slope will be lower than in other conditions). Here, we will report the most important findings, but the complete analysis scripts and final model results can be found at <https://osf.io/w38qg/>.

Results

Estimates. Figure 2A shows the average estimates for each duration and estimation and calibration condition. Model comparison showed that adding duration as a continuous fixed factor improved the basic model that included estimate as the dependent variable and subject as a random factor ($\chi^2(1) = 7647.90, p < .001$), indicating that, overall, estimates increased with the presented duration. In addition, estimation method and its interaction with duration improved the model fit ($\chi^2(2) = 63.33, p < .001$ and $\chi^2(2) = 33.89, p < .001$, respectively), showing that the intercept and slope of the estimates differed between estimation methods. Post-hoc contrasts showed that estimates (at the middle interval of 2.5 s) were longer for verbal estimates than for line estimates ($\beta = 0.13, p = .002$) and motor reproductions ($\beta = 0.10, p = .005$). In addition, the slopes of line estimates and motor reproductions were smaller than for verbal estimates ($\beta = -0.08, p < .001$ and $\beta = -0.09, p < .001$), suggesting a larger central tendency effect for line estimate and motor reproductions. There was no evidence for intercept or slope differences between the line estimates and motor reproduction condition ($ps > .666$). Adding calibration condition as a fixed factor did not improve the model fit ($\chi^2(1) = 0.85, p = .361$). However, we found a significant three-way interaction between duration, estimation method and calibration condition ($\chi^2(2) = 6.13, p = .047$). Post-hoc contrasts showed that the slope difference between the calibration condition was higher for motor reproductions than for line estimates ($\beta = 0.08, p = .031$).

Absolute error. Figure 2B shows the average absolute error for the different presented durations and experimental conditions. Presented duration improved the model fit ($\chi^2(1) = 467.25, p < .001$), indicating that overall the absolute error increased with duration. Adding estimation method as a fixed factor also improved the model ($\chi^2(2) = 128.70, p < .001$). Post-hoc contrasts showed that the error of verbal estimations was lower than for line estimations ($\beta = -0.14, p < .001$) and motor reproduction ($\beta = -0.07, p = .004$), and lower for motor reproductions than for line estimations ($\beta = -0.07, p < .001$). Model comparison showed that the interaction between presented duration and estimation method also improved the fit ($\chi^2(2) = 7.71, p = .021$). Post-hoc contrasts revealed a larger slope for motor reproductions compared to verbal estimations ($\beta = 0.03, p = .033$). Adding calibration condition did not improve the model fit ($\chi^2(1) = 1.01, p = .314$).

Interestingly, Figure 2B shows that the error in the verbal estimations systematically diverged from a linear pattern: visual inspection suggests that it was lower for integer durations (1, 2, 3 and 4 s) than for the durations in between (1.5, 2.5, and 3.5 s). Post-hoc, we tested this notion by adding a dichotomous fixed factor indicating whether a du-

ration was an integer to a model predicting the absolute error in the verbal estimation condition. Duration, calibration version and their interaction were also included as fixed factors. We found that this dichotomous fixed factor improved the model significantly ($\chi^2(1) = 71.93, p < .001$), indicating that the error was indeed lower for rounded integers. This was not the case for the line estimations ($\chi^2(1) = 3.46, p = .063$) and the motor reproductions ($\chi^2(1) = 0.83, p = .363$).

Coefficient of Variation (CV). We calculated the CV per participant and presented duration as the standard deviation divided by the average estimate. Figure 2C shows the average CV for every presented duration for the different estimation and calibration conditions. Presented duration improved the model significantly ($\chi^2(1) = 128.28, p < .001$), showing that - overall - the CV was smaller for longer durations. We found no evidence that this negative slope differed between estimation conditions ($\chi^2(2) = 4.26, p = .119$). However, we found that the intercept (at 2.5 s) did differ between estimation conditions ($\chi^2(2) = 50.75, p < .001$): In line with the absolute error, the CV was larger for line estimates and motor reproductions compared to the verbal estimates ($\beta = 0.06, p < .001$ and $\beta = 0.03, p = .007$, respectively) and larger for line estimates compared to motor reproductions ($\beta = 0.03, p = .004$). We found no evidence for a difference between the calibration conditions ($\chi^2(1) = 1.27, p = .260$).

End of the line effects. We expected that the calibration conditions would mostly affect the participants' tendency to not respond close to the end of the line. In this case, we would expect that calibration most strongly influences estimates of longer intervals, and that this effect would be most pronounced in line estimates. To test this hypothesis, we investigated the influence of calibration on the accuracy and precision of the longest interval (i.e., 4 s). An LMM predicting these estimates showed that they differed between estimation methods ($\chi^2(2) = 17.89, p < .001$). However, we found no evidence that calibration improved the overall estimates ($\chi^2(1) = 3.06, p = .080$), or that the calibration effect differed between estimation methods ($\chi^2(2) = 0.27, p = .874$). Looking at the precision, we also found that the absolute error at the longest interval differed between estimation methods ($\chi^2(2) = 26.24, p < .001$), and that the error was higher in the calibrated condition ($\chi^2(1) = 4.61, p = .032$). Although the visual inspection of Figure 2B suggests that the effect of calibration condition was larger for the timeline estimations compared to the other methods, we found no evidence that this effect differed between estimation methods ($\chi^2(2) = 4.29, p = .117$). The CV showed a similar pattern: it differed between estimation methods ($\chi^2(2) = 6.12, p = .047$) and was higher in the calibrated condition ($\chi^2(1) = 8.17, p = .004$), but there was no evidence for a difference in the effect of calibration between estimation methods ($\chi^2(2) = 4.26, p = .119$). Overall, these results indicate that calibrating participants with longer durations did not improve the accuracy of the line, motor or verbal estimates, but did decrease their precision.

Sequential context effects. To test whether there were differences in sequential context effects between the esti-

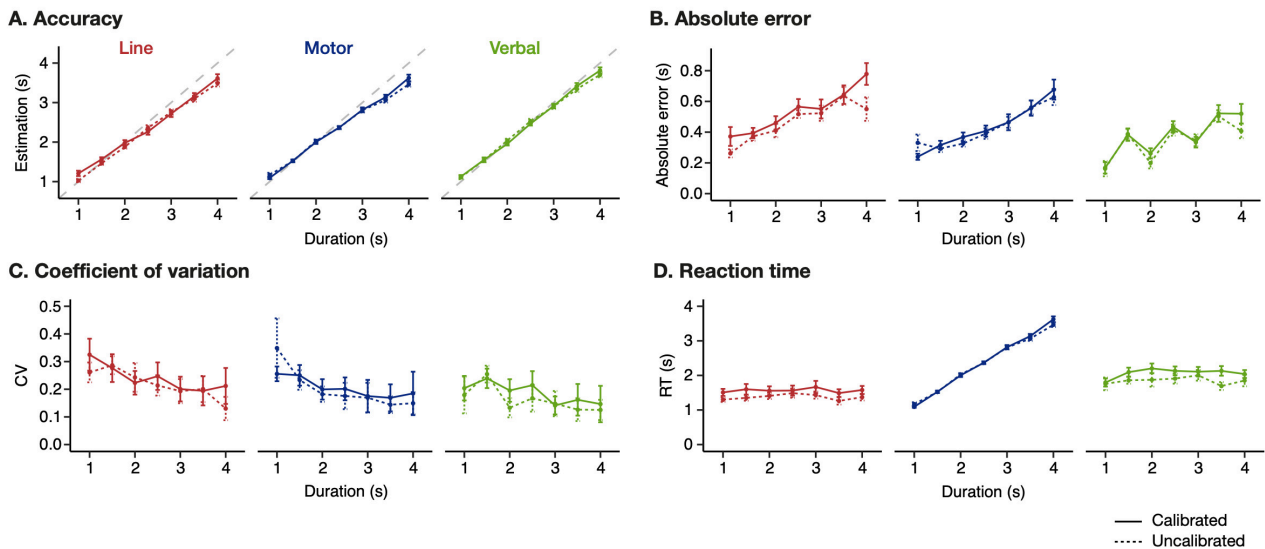


Figure 2. A) Average estimates for the timeline, motor and verbal conditions and calibration conditions in Experiment 1. The grey dashed line represents veridical performance. B) Average absolute error of the timeline, motor and verbal estimates and calibration conditions. C) Average CV of the timeline, motor and verbal estimates and calibration conditions. D) Average reaction times (RTs) of the timeline, motor and verbal estimates and calibration conditions. While the RTs are stable over durations for the timeline and verbal estimates, motor reproductions of course scale with the presented duration. In all figures, error bars represent the standard error of the mean.

mation methods, we tested the impact of previously presented durations. We started with the LMM predicting estimated duration including estimation condition, presented duration and their interaction as fixed factors. We gradually added previous presented durations (N-1, N-2, etc.) to the model as continuous fixed factors and tested whether they improved the model fit. We found that only the most recent previous trial (i.e., N-1) improved the model ($\chi^2(1) = 75.28$, $p < .001$), and that this factor differed between the estimation conditions ($\chi^2(2) = 7.07$, $p = .029$). Post-hoc contrasts showed that the effect of N-1 was larger for motor compared to verbal reproductions ($\beta = 0.04$, $p = .017$). There were no other differences ($ps > .239$).

Reaction time. Figure 2D shows the average reaction time (RT) for every estimation method and calibration condition. The model showed that the overall reaction time, and the change of RT with duration, differed between estimation methods ($\chi^2(2) = 731.77$, $p < .001$ and $\chi^2(2) = 848.88$, $p < .001$, respectively). Post-hoc contrasts showed that the RT at the 2.5 s interval intercept was higher for the verbal compared to the line estimation ($\beta = 0.49$, $p < .001$) and higher for the motor reproductions compared to the verbal ($\beta = 0.40$, $p < .001$) and line estimation ($\beta = 0.89$, $p < .001$). Because the motor reproductions increased with the presented duration, whereas reaction time for verbal and line estimations is independent of the presented duration, the slope was larger for the motor reproduction method ($\beta = 0.77$, $p < .001$ and $\beta = 0.79$, $p < .001$, respectively). There was no difference in slope between the verbal and line estimation ($p = .837$). Adding the interaction between estimation method and calibration condition improved the model

significantly ($\chi^2(2) = 14.40$, $p < .001$), but there were no significant differences in the final model ($ps > .328$).

Discussion

In Experiment 1, we compared three estimation methods in a simple interval estimation task. The results showed that the verbal estimates were overall more veridical than the motor and line estimates. We found no evidence for a difference in the accuracy of the motor and line estimates. When we look at precision of estimation, we found that the CV decreased with the presented duration. This is a violation of Weber's law, or the scalar property of time perception, which states that the CV should be constant over different durations (although violations are frequent in the timing literature: see Grondin, 2014). Comparing the absolute error and the CV between estimation methods, we found that verbal estimates were most precise. Notably, however, this precision depended on the specific presented duration: it was higher for rounded integers than for durations with a fractional part. In addition, motor reproductions were generally more precise than line estimates. Overall, these results suggest that there is no cost in accuracy to the potential spatial transformation required for line estimates, but there might be a small cost in precision, that is, the variability of the estimates. Note, however, that any difference between motor reproductions and line estimates, especially, may arise due to differences in the amount of motor noise rather than because of their underlying representation and translation into another dimension.

We expected that calibrating participants with a larger interval range and a longer corresponding timeline at the

start of the experiment would diminish the underestimation of longer durations. However, we found no evidence that this calibration increased the overall accuracy, or the accuracy of the longest duration. Instead, we found a small cost in the precision of the longest interval estimates. These results suggest that calibration did not improve the timeline estimates by diminishing a potential end of the line bias. Alternatively, the end of the line effects here (and in Damsma et al., 2018) could reflect a general pull towards the mean in which the estimate is biased towards previously presented durations. Indeed, sequential context effects were observed in all three conditions. Verbal estimates were less affected by the duration of the previous trial, which can be explained by their generally higher accuracy and precision. According to the Bayesian view of perception, it is optimal to rely more on prior experience in making an estimate when the current observation is less precise (Acerbi et al., 2012; Jazayeri & Shadlen, 2010).

When reproducing longer time intervals using a motor response, the duration of a trial scales with the interval to be reproduced. Line and verbal estimates, on the other hand, have the advantage of a stable response time of - in our experiment - around 1.5 and 2 s, respectively. Based on these results, we suggest that researchers can increase the number of trials in their experiment when testing longer intervals by using line or verbal estimates, and thereby increase statistical power.

Another potential advantage of timeline estimates over motor reproductions is that there is more time for a deliberate decision, compared to the 'one shot' approach of motor reproductions: in the latter case, participants are by definition unable to decrease their estimate at any point in time. This 'time asymmetry' might induce biases specific to motor reproduction, such as systematic under-reproduction (Riemer et al., 2012). In contrast, in the line estimation condition, participants can move the cursor freely to the left or right to decrease or increase their estimate. This could make timeline estimates more accurate in situations that are more complex than the reproduction of a single interval to which participants can fully direct their attention, such as when estimating an interval concurrently with other tasks (e.g., Brown, 1997, 2006; Zakay, 1993) or estimating multiple intervals (e.g., Brown & West, 1990; van Rijn & Taatgen, 2008). To test this notion, Experiment 2 consisted of a stream of stimuli that contained one target. Participants had to estimate both the target onset and the duration of the stream.

Experiment 2

Methods

Participants. Thirty-nine healthy adults (9 male, mean age 20.64 years) participated in exchange for course credits. None of the participants in Experiment 1 took part in Experiment 2. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen (identification number 17054-S-NE) was obtained before testing. Sample size was based on past research (e.g., Experiment 1; Damsma et al., 2018; Schlichting et al., 2018), no statistical *a priori* power analysis was conducted.

Experimental Design. Participants were asked to per-

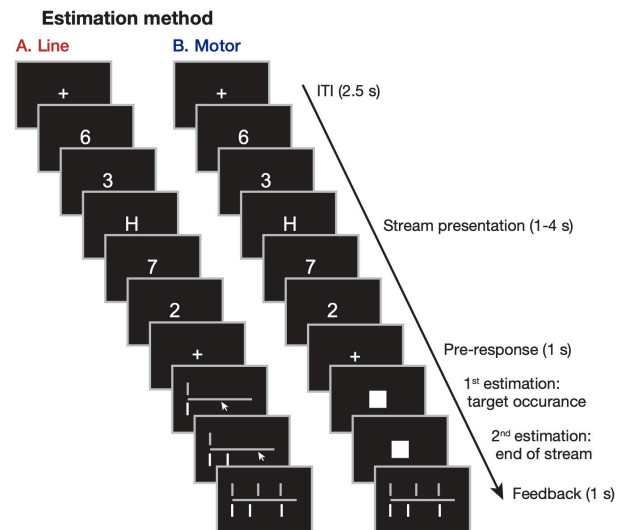


Figure 3. Trial procedure of Experiment 2. Participants were presented with a stream of numbers with one target letter. Their task was to estimate the interval from the beginning of the stream until the target onset, and also of the total duration of the stream by either A) clicking on a timeline (line estimates) or B) pressing a key at the estimated moments (motor reproduction). Feedback was presented at the end of each trial.

form a temporal estimation task using two different methods: motor reproductions and timeline estimations (Figure 3). At the beginning of the experiment, participants were explicitly instructed to not keep track of time by, for example, counting or feet tapping. Stimuli were displayed on a 1280 × 1024 CRT-based monitor screen (Iiyama Vision Master Pro 513) with a refresh rate of 100 Hz.

The interval was presented as a stream of numeric characters (1 to 9 characters in total) and one alphabetic character, the target (A, B, C, D, E, F, H, J, K, P, R, T, U, or V). The alphanumeric characters were presented in Arial with a font size of 16 pt. Within the stream alphanumeric characters were chosen randomly, while no two consecutive characters were the same. Participants were asked to estimate both the interval from stream onset to target onset as well as the duration of the total stream. There were six different total stream durations (4.75, 5.25, 5.75, 6.25, 6.75, and 7.25 s) and 11 positions where the target could occur from stream onset (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, and 6 s). Target onset was chosen completely random, that is, for some participants not all target positions may have occurred. Each alphanumeric character was presented for 0.25 s with 0.25 s between two successive characters, so that each stream consisted of 10 to 15 alphanumeric characters in total. The estimation methods were similar to Experiment 1, with the only difference that two responses were required. In the motor reproduction task, a first spacebar press corresponded to the time point of target occurrence, and a second spacebar press corresponded to the end of the stream. Similarly, a first mouse click on the timeline corresponded to target occurrence, and a second mouse click to the end of the stream in

the timeline estimation task. Participants received immediate feedback similar to the feedback in Experiment 1, with two additional bars corresponding to the veridical and estimated target occurrence (or interval to target onset).

The experiment was divided into two blocks (i.e., one block for each estimation method) of 60 experimental trials (i.e., ten trials per duration) each. Order of blocks, and thus estimation methods, was counterbalanced between participants. Before the start of each block, participants received instructions about the estimation method to be used in the upcoming block, and they performed 12 practice trials in order to get accustomed to the timeline and the estimation method before the start of the experimental trials. The order of trials was the same in each block but varied between participants.

As in Experiment 1, half of the participants performed a calibrated version of the estimation tasks. In the calibrated version the training-trials consisted also of longer overall intervals than those in the test trials (3.25, 6.25, and 9.75 s), while in the uncalibrated version training trials were chosen up to the longest duration of test trials (2.25, 4.75, and 7.25 s). The experiment files can be found at <https://osf.io/w38qg/>.

Analysis. All estimates shorter than 0.2 s and longer than 11 s (1.95% of the data) were excluded from analysis. The longest durations of the target estimates (5, 5.5 and 6 s) were also excluded from analysis, because there were on average less than 4 trials per condition per participant, leading to unreliable calculations of the error and CV measures. The analysis procedure was similar to Experiment 1. Target and stream estimates were analyzed separately. The durations were centered at 2.75 s and 6 s for target and stream estimations, respectively. All categorical fixed factors were recoded using effect coding (-0.5 and 0.5), to facilitate the interpretation of main effects when interactions are included in the model. In the current experiment, there were two estimation methods (motor reproduction and timeline estimation) instead of three methods in Experiment 1. Therefore, instead of post-hoc contrast results, we will report the β -coefficient and t -value of factors in the final LMM, as they are a direct representation of the difference between the estimation methods. The analysis scripts and results can be found at <https://osf.io/w38qg/>.

Results

Estimates. *Interval to target onset.* Figure 4A shows the average estimates of the interval between stream onset and target onset for the different conditions. We found that, overall, target estimates increased with the presented onset ($\chi^2(1) = 2105.36, p < .001; \beta = 0.66, t = 19.54$). In addition, motor reproductions were shorter than line estimates ($\chi^2(1) = 26.31, p < .001; \beta = -0.14, t = -2.94$). No other fixed effects reached significance.

Total stream duration. The stream estimates also increased with the presented duration ($\chi^2(1) = 792.99, p < .001; \beta = 0.57, t = 16.87$), but here the slope was steeper for motor reproductions than for line estimates ($\chi^2(1) = 7.13, p = .008; \beta = 0.09, t = 2.55$). Model comparison showed a stronger effect of calibration for the line estimates com-

pared to the motor reproduction methods ($\chi^2(1) = 24.85, p < .001$), although this effect did not reach significance after including random slopes in the final model ($\beta = -0.35, t = -1.68, p = .101$). In addition, the effect of calibration on the *slope* was larger for the line estimates compared to the motor reproductions ($\chi^2(1) = 4.44, p = .035; \beta = -0.17, t = -2.31$). Overall, these results suggest that stream estimates were more veridical for the motor compared to the line condition, but that calibration with a longer timeline decreased this difference.

Absolute error. *Interval to target onset.* Figure 4B shows the average absolute error for the estimations of different durations for each condition. The LMM showed that the absolute error of target estimates increased with duration ($\chi^2(1) = 5.37, p = .020; \beta = 0.03, t = 2.57$). Model comparison suggested that there was a difference between line estimates and motor reproductions ($\chi^2(1) = 4.60, p = .032$), but this effect was not significant after including random slopes ($\beta = -0.06, t = -1.12, p = .271$). Overall, calibration condition did not affect the absolute error.

Total stream duration. In line with the absolute error of the target estimates, the error of the stream estimates increased with duration ($\chi^2(1) = 45.02, p < .001; \beta = 0.15, t = 9.60$). In addition, model comparison showed an interaction effect of estimation method and calibration condition, but this effect did not remain significant in the final model ($\chi^2(1) = 5.69, p = .017; \beta = 0.17, t = 0.96$). However, the slope difference between the calibration conditions was larger for motor compared to line estimates ($\chi^2(1) = 4.64, p = .031; \beta = 0.18, t = 3.00$). The final model also revealed a steeper slope for the motor compared to the line condition ($\beta = 0.07, t = 2.34$).

Coefficient of Variation (CV). *Interval to target onset.* Figure 4C shows the CV for the different duration and conditions. We found that the CV decreased with duration ($\chi^2(1) = 75.60, p < .001; \beta = -0.04, t = -8.98$). We found no differences between the estimation methods or the calibration conditions.

Total stream duration. In contrast to the target estimates, the CV of the stream estimates did not change with duration ($\chi^2(1) = 0.44, p = .509$). However, the slope was more positive for motor compared to line estimates ($\chi^2(1) = 4.85, p = .028; \beta = 0.02, t = 2.21$).

Sequential context effects. *Interval to target onset.* We started with the LMM established to predict the target estimates (including duration and estimation method as fixed factors). We then sequentially added previous target and stream durations. We found that target estimates were significantly influenced by target estimates in the previous trial (i.e., N-1; $\chi^2(1) = 23.16, p < .001; \beta = 0.05, t = 4.82$). This effect did not differ between the motor and line estimates. There was no significant effect of N-2 ($\chi^2(1) = 2.65, p = .104$). We also tested whether the stream estimates in the current trial influenced the target estimates, but there was no evidence that this was the case ($\chi^2(1) = 0.89, p = .345$).

Total stream duration. The estimates of the stream durations were influenced by the stream duration in the previous trial ($\chi^2(1) = 38.24, p < .001; \beta = 0.14, t = 7.33$). This ef-

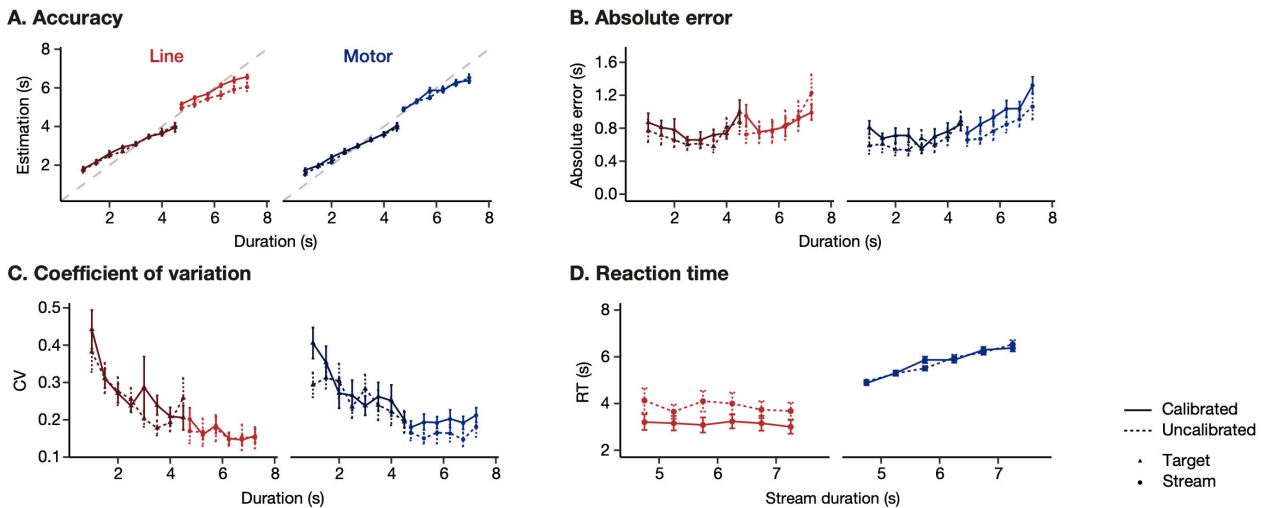


Figure 4. A) Average interval-to-target and total stream duration estimates for the timeline and motor conditions and calibration conditions in Experiment 2. The grey dashed line represents veridical performance. B) Average absolute error of the target and stream estimates for the timeline and motor conditions and calibration conditions. C) Average CV of the target and stream estimates for the timeline and motor conditions and calibration conditions. D) Average response time (RT) of the stream estimation for the timeline and motor conditions and calibration conditions. In all figures, error bars represent the standard error of the mean.

fect did not differ statistically between estimation methods ($\chi^2(1) = 2.89, p = .089$). The stream estimates were also influenced by target onset ($\chi^2(1) = 27.55, p < .001$; $\beta = 0.66, t = 5.14$) and target onset in the previous trial ($\chi^2(1) = 9.59, p = .002$; $\beta = -0.04, t = -3.18$). The latter effect was stronger for line estimates compared to motor reproductions ($\chi^2(1) = 8.23, p = .004$; $\beta = 0.06, t = 2.59$).

Reaction time. Figure 4D shows the average response times. We found that duration, estimation method and their interaction improved the model fit ($\chi^2(1) = 47.82, p < .001$, $\chi^2(1) = 1282.98, p < .001$ and $\chi^2(1) = 5.95, p = .015$, respectively). In line with Figure 4D, the final model showed that RTs were higher ($\beta = 0.91, t = 3.21$), and the increase of RTs with duration was larger ($\beta = 0.71, t = 11.63$), for motor reproductions compared to line estimates. There was no overall effect of calibration condition ($\chi^2(1) = 2.33, p = .013$). Although the interaction between estimation method and calibration condition improved the model ($\chi^2(1) = 43.98, p < .001$), this fixed effect was not significant in the eventual model including random slopes ($\beta = 0.76, t = 1.47$).

Discussion

In Experiment 2, participants were asked to reproduce the interval between the onset of an alphanumeric stream and a target letter in the stream as well as the end of the stream. The results suggest that motor reproductions had a slightly more veridical slope than line estimates, but only for the stream estimates. This can be explained by a tendency by the participants to avoid the end of the line, leading to a relative compression of stream estimates, which need to be placed towards the end of the scale. This notion

is also reflected in the effect of calibration: while calibration had no effect on motor reproductions, it improved the average accuracy of the stream estimates in the line condition. In line with Experiment 1, the CV decreased with duration, violating the scalar property. The overall precision of the responses of the motor reproductions and line estimates was similar, however, the variability increased more with the presented duration for motor reproductions. As in Experiment 1, we found that reaction times were stable over the different test durations and lower overall in the line condition.

We again found that previously perceived target or stream durations influenced target and stream estimates in the current trial, and there was no difference between estimation methods. There was also an effect of target onset on stream estimates, in that the later the target appeared, the longer the stream was estimated. One explanation is that participants use a sort of relative timing: if the target occurred relatively late, the duration of the stream was probably longer (see also van Rijn & Taatgen, 2008). In the line estimation condition, another explanation of this finding is that participants tend to keep their distance from the target estimates when making the second estimates on the stream duration, an effect that might be similar to the bias of avoiding the end of the scale. Thus, if the target occurred relatively late, the stream duration estimate will be shifted to having occurred later (see also Damsma et al., 2018, who show that estimates of the timing of targets in an attentional blink paradigm are not independent of each other). Interestingly, we found a difference between estimation methods in the effect of the previous target onset on stream estimates. This effect may have been more prevalent in the line estimation condition because of the strong visual

representation in the line compared to the motor condition. Not only were participants able to see their target estimate in the line condition, but it was also potentially easier to incorporate the feedback of the previous trial because it was visualized in the exact same way as participants gave their estimates. Because of the increased task complexity in Experiment 2, the visual representation might have been taken more into account as compared to Experiment 1.

General discussion

In the current study, we compared the accuracy and precision of interval estimations using a visual analogue scale (or, a timeline) to non-spatial estimation methods (motor reproductions in Experiment 1 and 2 and verbal estimations in Experiment 1). If, regardless of estimation method, temporal estimates undergo the same or similar transformations, we expected to find no differences between the different estimation methods. If, on the other hand, a mental transformation from time to space is required, we would expect costs in accuracy and precision in the timeline estimates. In Experiment 1, we found similar accuracy for line estimates and motor reproductions, whereas the precision was higher for motor estimates. Verbal estimates seemed to lead to the most accurate and precise estimates. However, the pattern we found in absolute errors suggests that this estimation method comes with its own unique problems that we discuss further below. In the more complex paradigm of Experiment 2 we found that estimates were slightly more accurate for motor reproductions compared to timeline estimates, while the precision was similar.

Taken together, these results suggest that both motor reproduction and timeline estimation can be reliably used to measure subjective timing. This could potentially indicate that space and time have a similar neural representation (e.g., Walsh, 2003, 2015). For example, in the episodic memories of everyday events, hippocampal neurons might encode both space and time information (Buzsáki, 2019; Buzsáki & Llinás, 2017). Whether a common representation of space and time is inherent in low-level accumulation or only occurs at later stages in the representation or decision process (e.g., Anobile et al., 2018) remains an open question. Alternatively, it is possible that the transformation of duration into space has only a relatively minor cost, roughly equivalent to the effect of noise introduced by manual reproduction. Time may be represented in a sufficiently abstract way to make transformations to any other representational form effortless and equally accurate; or time is omnipresent in any neural process and readily available as input for other cognitive processes (Hass & Durstewitz, 2014).

In either case, however, it is important to note that both motor reproductions and line estimates might come with their own respective sources of noise. Models of the interval reproduction task often assume that variance is introduced in the motor response, independently of the variance in interval perception (Acerbi et al., 2012; Wearden, 2003). This additional motor noise would presumably decrease the precision of the estimates, relative to methods that do not require a motor response or methods that use alternative, less noisy motor responses such as eye movements. An addi-

tional factor that could lower precision of motor reproductions is the previously discussed ‘one-shot’ approach. Finally, given that motor response might play an important role in the effect of previous durations on the current estimate (Roach et al., 2017), motor reproductions could show a larger pull towards the mean, decreasing accuracy. However, we found no difference in the effect of the previous trial on the current estimate between motor and line estimates. This is in line with recent evidence from our lab, suggesting that previous durations already affect *perception* itself (Damsma et al., 2020). Timeline estimation also has its own unique sources of variance. Participants first have to learn how exactly time translates into space when using a specific timeline. This means that even if time would be represented spatially, a source of noise in line estimates could be that participants have to scale their spatial representation of time before giving an estimate. Because it is difficult to disentangle these sources of noise from noise in the representation of time, the differences in accuracy and precision between the estimation methods do not allow arguing in favor or against the idea of a spatial time or general magnitude representation.

In Experiment 1, we found that verbal estimates were more accurate and precise than timeline estimates and motor reproductions, implying that verbal estimates are superior to other estimation methods. However, participants were encouraged to express their subjective estimate in familiar terms (in this study: seconds). This familiarity might come at a cost: We found that the verbal estimates displayed an inconsistent pattern of precision, in which rounded integer intervals were estimated with higher precision than non-integer intervals. This pattern can be explained in three ways: 1) the emphasis on ‘seconds’ might lead participants to think about time in terms of these pre-learned units, 2) the method might have encouraged participants to count (although they were explicitly instructed not to count), and 3) the method of report might have encouraged some participants to round their estimate to the nearest integer, without using the fractional part. Regardless of the origin of the precision pattern, the results indicate that verbal estimates might encourage participants to think about time in a less ‘linear’ and a more ‘categorical’ way. Indeed, this is in line with the idea that verbal estimates are “contaminated by linguistic and semantic tags associated with traditional units of time perception” (Hancock & Block, 2012). These hypotheses imply that verbal estimates might be less accurate or precise when, for example, a range of sub-second intervals is reproduced. Future studies might test this idea by comparing estimation methods in different interval ranges.

In both experiments, we found that the coefficient of variation decreased with duration. This is a violation of the scalar property, which states that the CV should be constant over estimated durations. While the scalar property holds true for certain time scales and paradigms, it has to be noted that violations of the scalar property also have been observed commonly in timing tasks (for a review, see Grondin, 2014). Indeed, several studies have shown a decrease in CV similar to the effect reported here (e.g., Laje et al., 2011; Lewis & Miall, 2009). A potential explanation for the decreasing CV is that the variance observed in the

task has both time-dependent (variability associated with the timing mechanism) and time-independent components (such as motor variability), as noted above. Laje et al. (2011) showed that the decrease in CV could be captured by a generalized form of Weber's law, in which these components are explicitly modeled. Time-independent variance naturally depend on the nature of the estimation paradigm (for example, the extent to which the estimation method depends on motricity), explaining the differences in CV between estimation conditions in Experiment 1.

The differences in the results of Experiment 1 and 2 might be due to the different paradigms. First, participants had to reproduce two intervals in Experiment 2 (i.e., target onset and stream offset), and only a single interval in Experiment 1. This makes the task more difficult, which results in a higher absolute error in Experiment 2 (see also Brown et al., 1992; Brown & West, 1990). In addition, the estimates of the first and second interval might not be completely independent: the results show that there is an intercept difference, accompanied by a 'local' pull towards the mean, dependent on whether the first or the second interval is reproduced (see also Damsma et al., 2018). These dependencies might be stronger when they are visually represented on a timeline compared to motor reproductions. Second, longer intervals were presented in Experiment 2, which would also decrease the precision, in line with the scalar property.

One explanation for the lower accuracy in the timeline estimates in Experiment 2 is an increased response bias (i.e., reluctance to use the end of the scale), because the timeline offers a more explicit physical range compared to motor reproductions. If this is the case, we expected that estimates would be more accurate when the interval range is artificially increased in pre-experiment calibration trials. Indeed, the results of Experiment 2 showed that calibrating participants with a larger range increased the accuracy for longer intervals (i.e., the stream duration estimates), with similar precision. In Experiment 1, the calibration neither improved the overall accuracy, nor the accuracy of the longer intervals. Overall, these results suggest that the range of the timeline should be taken into account, as a range that is larger than the actual test durations might reduce the response bias for longer intervals. In the current study, the resolution of the timeline was identical in the calibrated and non-calibrated condition (1 pixel on screen corresponded to 0.01 s). Future studies could test whether this property of the timeline affects accuracy and precision of estimates, especially if the range of test durations is much larger than in the current study. Additionally, participants in our experiments received feedback about their accuracy on a line in every estimation condition, to keep the conditions as similar as possible. This way of presenting feedback could potentially bias participants towards a spatial representation of time, or benefit performance for motor or line estimates differentially. Future studies might test this notion by removing the feedback or varying the feedback modality.

Overall, the results show that each estimation method comes with its own unique advantages and drawbacks. Line estimations offer the advantage of a stable response time, which can allow the researcher to increase the number of

trials in supra-second interval estimation experiments (i.e., using intervals longer than ~1.5 s). However, compared to motor reproductions, there might be a small cost in accuracy, potentially because of a required spatial transformation. This difference might be overcome by calibrating participants with a suitable interval range. Motor reproductions offer an intuitive estimation method, but the response times scale linearly with the presented intervals. In addition, it is difficult to disentangle the precision of the actual temporal estimate from motor inaccuracies (Droit-Volet, 2010; Hallez et al., 2019). In Experiment 1, we showed that verbal estimates are more accurate and precise than line estimates and motor reproductions. However, the precision of verbal estimates depends on whether the interval is a whole integer, indicating a bias towards familiar whole second units. Although future research should further investigate the reliability of estimation methods in different timing experiments, the current study can point timing researchers to a more optimal estimation method given their specific paradigm.

Contributions

AD, NS, HvR, and WR conceived the study. AD and NS programmed the experimental paradigm, collected and analyzed the data. AD and NS wrote the initial manuscript. HvR and WR reviewed and critically revised the manuscript. All authors approved the submitted version for publication.

Acknowledgements

We thank Brendan Kanter for his help with data acquisition.

Funding Information

This research has been partially supported by the EU Horizon 2020 FET Proactive grant TimeStorm – Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence (grant no. 641100). Hedderik van Rijn is additionally funded by the research program "Interval Timing in the Real World: A functional, computational and neuroscience approach" (project number 453-16-005) financed by the Netherlands Organization for Scientific Research (NWO). The funding agencies were not involved in the design of the study; the analysis and interpretation of the data; writing of the report; and in the decision to submit the article for publication.

Competing Interests

The authors have no competing interests to declare.

Data Accessibility Statement

All presentation materials, participant data, and analysis scripts can be found online at [10.17605/OSF.IO/W38QG](https://doi.org/10.17605/OSF.IO/W38QG).

Submitted: May 21, 2019 PDT, Accepted: February 18, 2021 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

- Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal Representations of Temporal Statistics and Feedback Calibrate Motor-Sensory Interval Timing. *PLoS Computational Biology*, 8(11). <https://doi.org/10.1371/journal.pcbi.1002771>
- Anobile, G., Burr, D. C., Iaia, M., Marinelli, C., Angelelli, P., & Turi, M. (2018). Independent adaptation mechanisms for numerosity and size perception provide evidence against a common sense of magnitude. *Scientific Reports*, 8(1), 1–12. <https://doi.org/10.1038/s41598-018-31893-6>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bender, A., & Beller, S. (2014). Mapping spatial frames of reference onto time: a review of theoretical accounts and empirical findings. *Cognition*, 132(3), 342–382. <https://doi.org/10.1016/j.cognition.2014.03.016>
- Block, R. A., Grondin, S., & Zakay, D. (2018). Prospective and retrospective timing processes: Theories, methods, and findings. In A. Vatakis, F. Balci, M. Di Luca, & Á. Correa (Eds.), *Timing and Time Perception: Procedures, Measures, and Applications* (pp. 32–51). Brill. https://doi.org/10.1163/9789004280205_003
- Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews*, 36(10), 2257–2273. <https://doi.org/10.1016/j.neubiorev.2012.08.007>
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1–22. <https://doi.org/10.1006/cogp.2001.0748>
- Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, 118(1), 123–129. <https://doi.org/10.1016/j.cognition.2010.09.010>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59(7), 1118–1140. <https://doi.org/10.3758/BF03205526>
- Brown, S. W. (2006). Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, 34(7), 1464–1471. <https://doi.org/10.3758/BF03195911>
- Brown, S. W., Stubbs, D. A., & West, A. N. (1992). Attention, multiple timing, and psychophysical scaling of temporal judgments. In *Time, Action and Cognition* (pp. 129–140). Springer Netherlands. https://doi.org/10.1007/978-94-017-3536-0_15
- Brown, S. W., & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica*, 75(2), 103–121. [https://doi.org/10.1016/0001-6918\(90\)90081-P](https://doi.org/10.1016/0001-6918(90)90081-P)
- Buetti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1525), 1831–1840. <https://doi.org/10.1098/rstb.2009.0028>
- Buzsáki, G. (2019). The Brain from Inside Out. In *The Brain from Inside Out*. Oxford University Press, USA. <https://doi.org/10.1093/oso/9780190905385.001.0001>
- Buzsáki, G., & Llinás, R. (2017). Space and time in the brain. *Science*, 358(6362), 482–485. <https://doi.org/10.1126/science.aan8869>
- Cai, Z. G., & Connell, L. (2016). On magnitudes in memory: An internal clock account of space-time interaction. *Acta Psychologica*, 168, 1–11. <https://doi.org/10.1016/j.actpsy.2016.04.003>
- Cai, Z. G., Wang, R., Shen, M., & Speekenbrink, M. (2018). Cross-dimensional magnitude interactions arise from memory interference. *Cognitive Psychology*, 106, 21–42. <https://doi.org/10.1016/j.cogpsych.2018.08.001>
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, 106(2), 579–593. <https://doi.org/10.1016/j.cognition.2007.03.004>
- Conson, M., Cinque, F., Barbarulo, A. M., & Trojano, L. (2008). A common processing system for duration, order and spatial information: Evidence from a time estimation task. *Experimental Brain Research*, 187(2), 267–274. <https://doi.org/10.1007/s00221-008-1300-5>

- Coull, J. T., & Droit-Volet, S. (2018). Explicit understanding of duration develops implicitly through action. *Trends in Cognitive Sciences*, 22(10), 923–937. <https://doi.org/10.1016/j.TICS.2018.07.011>
- Coull, J. T., Johnson, K. A., & Droit-Volet, S. (2018). A mental timeline for duration from the age of 5 years old. *Frontiers in Psychology*, 9, 1155. <https://doi.org/10.3389/fpsyg.2018.01155>
- Damsma, A., Schlichting, N., & Rijn, H. van. (2020). Temporal context actively shapes EEG signatures of time perception. *BioRxiv*, 2020.03.17.995704. <https://doi.org/10.1101/2020.03.17.995704>
- Damsma, A., van der Mij, R., & van Rijn, H. (2018). Neural markers of memory consolidation do not predict temporal estimates of encoded items. *Neuropsychologia*, 117, 36–45. <https://doi.org/10.1016/j.neuropsychologia.2018.04.039>
- Dormal, V., Dormal, G., Joassin, F., & Pesenti, M. (2012). A common right fronto-parietal network for numerosity and duration processing: An fMRI study. *Human Brain Mapping*, 33(6), 1490–1501. <https://doi.org/10.1002/hbm.21300>
- Droit-Volet, S. (2010). Stop using time reproduction tasks in a comparative perspective without further analyses of the role of the motor response: The example of children. *European Journal of Cognitive Psychology*, 22(1), 130–148. <https://doi.org/10.1080/09541440902738900>
- Fabbri, M., Cancellieri, J., & Natale, V. (2012). The A Theory Of Magnitude (ATOM) model in temporal perception and reproduction tasks. *Acta Psychologica*, 139(1), 111–123. <https://doi.org/10.1016/j.actpsy.2011.09.006>
- Fabbri, M., Cellini, N., Martoni, M., Tonetti, L., & Natale, V. (2013). The mechanisms of space-time association: Comparing motor and perceptual contributions in time reproduction. *Cognitive Science*, 37(7), 1228–1250. <https://doi.org/10.1111/cogs.12038>
- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, 34(8), 1430–1451. <https://doi.org/10.1111/j.1551-6709.2010.01105.x>
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, 54(3), 383–394. <https://doi.org/10.3758/BF03205274>
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72(3), 561–582. <https://doi.org/10.3758/APP.72.3.561>
- Grondin, S. (2014). About the (Non)scalar property for time perception. *Advances in Experimental Medicine and Biology*, 829, 17–32. https://doi.org/10.1007/978-1-4939-1782-2_2
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4), 382–385. <https://doi.org/10.1038/nn827>
- Hallez, Q., Damsma, A., Rhodes, D., van Rijn, H., & Droit-Volet, S. (2019). The dynamic effect of context on interval timing in children and adults. *Acta Psychologica*, 192, 87–93. <https://doi.org/10.1016/j.actpsy.2018.10.004>
- Hancock, & Block. (2012). The Psychology of time: A view backward and forward. *The American Journal of Psychology*, 125(3), 267. <https://doi.org/10.5406/amerjpsyc.125.3.0267>
- Hass, J., & Durstewitz, D. (2014). Neurocomputational models of time perception. In *Advances in Experimental Medicine and Biology* (Vol. 829, pp. 49–71). https://doi.org/10.1007/978-1-4939-1782-2_4
- Hayashi, M. J., Kanai, R., Tanabe, H. C., Yoshida, Y., Carlson, S., Walsh, V., & Sadato, N. (2013). Interaction of numerosity and time in prefrontal and parietal cortex. *Journal of Neuroscience*, 33(3), 883–893. <https://doi.org/10.1523/JNEUROSCI.6257-11.2013>
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., & Hothorn, M. T. (2017). Package “multcomp.” See <https://cran.r-project.org/web/packages/multcomp/index.html>
- Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex*, 44(4), 454–461. <https://doi.org/10.1016/j.cortex.2007.08.010>
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8), 1020–1026. <https://doi.org/10.1038/nn.2590>
- Laje, R., Cheng, K., & Buonomano, D. (2011). Learning of temporal motor patterns: An analysis of continuous versus reset timing. *Frontiers in Integrative Neuroscience*, 5. <https://doi.org/10.3389/fnint.2011.00061>

- Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: Milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1897–1905. <https://doi.org/10.1098/rstb.2009.0020>
- Loeffler, J., Cañal-Bruland, R., Schroeger, A., Walter Tolentino-Castro, J., & Raab, M. (2018). Interrelations between temporal and spatial cognition: The role of modality-specific processing. *Frontiers in Psychology*, 9, 2609. <https://doi.org/10.3389/fpsyg.2018.02609>
- Magnani, B., & Musetti, A. (2017). Innate and cultural spatial time: A developmental perspective. *Frontiers in Human Neuroscience*, 11, 215. <https://doi.org/10.3389/fnhum.2017.00215>
- Mioni, G. (2018). Methodological issues in the study of prospective timing. In A. Vatakis, F. Balci, M. Di Luca, & Á. Correa (Eds.), *Timing and Time Perception: Procedures, Measures, and Applications* (pp. 79–97). Brill. https://doi.org/10.1163/9789004280205_005
- Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, 17(5), 220–229. <https://doi.org/10.1016/j.tics.2013.03.008>
- R Core Team, D. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Riemer, M., Diersch, N., Bublatzky, F., & Wolbers, T. (2016). Space, time, and numbers in the right posterior parietal cortex: Differences between response code associations and congruency effects. *NeuroImage*, 129, 72–79. <https://doi.org/10.1016/j.neuroimage.2016.01.030>
- Riemer, M., Trojan, J., Kleinböhl, D., & Hölzl, R. (2012). A “view from nowhen” on time perception experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 38(5), 1118–1124. <https://doi.org/10.1037/a0027073>
- Roach, N. W., McGraw, P., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences of the United States of America*, 114(2), 412–417. <https://doi.org/10.1073/pnas.1610706114>
- Roseboom, W., Fountas, Z., Nikiforou, K., Bhowmik, D., Shanahan, M., & Seth, A. K. (2019). Activity in perceptual classification networks as a basis for human subjective time perception. *Nature Communications*, 10(1), 1–9. <https://doi.org/10.1038/s41467-018-08194-7>
- Schlichting, N., Damsma, A., Aksoy, E. E., Wächter, M., Asfour, T., & van Rijn, H. (2018). Temporal Context Influences the Perceived Duration of Everyday Actions: Assessing the Ecological Validity of Lab-Based Timing Phenomena. *Journal of Cognition*, 2(1). <https://doi.org/10.5334/joc.4>
- Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial–temporal association of response codes: Understanding the cognitive representations of time. *Cognition*, 107(2), 501–527. <https://doi.org/10.1016/j.cognition.2007.10.011>
- van Rijn, H. (2016). Accounting for memory mechanisms in interval timing: A review. *Current Opinion in Behavioral Sciences*, 8, 245–249. <https://doi.org/10.1016/j.cobeha.2016.02.016>
- van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica*, 129(3), 365–375. <https://doi.org/10.1016/j.actpsy.2008.09.002>
- Vicario, C. M., Pecoraro, P., Turriziani, P., Koch, G., Caltagirone, C., & Oliveri, M. (2008). Relativistic compression and expansion of experiential time in the left and right space. *PLoS ONE*, 3(3), e1716. <https://doi.org/10.1371/journal.pone.0001716>
- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488. <https://doi.org/10.1016/j.tics.2003.09.002>
- Walsh, V. (2015). A Theory of Magnitude: The parts that sum to number. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition* (pp. 553–565). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642342.013.64>
- Wearden, J. H. (2003). Chapter 2: Applying the scalar timing model to human time psychology: Progress and challenges. In *Time and mind II: information processing perspectives* (pp. 21–39).
- Wearden, J. H. (2016). *The psychology of time perception*. Palgrave Macmillan UK. <https://doi.org/10.1057/978-1-137-40883-9>
- Wearden, J. H., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *Quarterly Journal of Experimental Psychology*, 59(10), 1709–1724. <https://doi.org/10.1080/017470210500314729>
- Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review*, 15(2), 426–430. <https://doi.org/10.3758/PBR.15.2.426>

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, 7(10), 2.1-5. <https://doi.org/10.1167/7.10.2.Introduction>

Zakay, D. (1993). Time estimation methods—Do they influence prospective duration estimates? *Perception*, 22(1), 91–101. <https://doi.org/10.1068/p220091>